



... children, their world, their education

PRIMARY REVIEW
RESEARCH BRIEFINGS
3/4

THE QUALITY OF LEARNING: ASSESSMENT ALTERNATIVES FOR PRIMARY EDUCATION

Wynne Harlen
University of Bristol

This briefing draws on Primary Review Research Report 3/4 *The Quality of Learning: assessment alternatives for primary education*, by Wynne Harlen. The report was commissioned to review representative published research on assessment approaches and issues, with a view to encouraging debate about the merits of the current English system and consideration of well-founded alternatives. Discussion of assessment tends to be dominated by the KS1/KS2 National Tests: Report 3/4 opens up a wider range of possibilities. **The full report, including details of sources consulted, is available at www.primaryreview.org.uk.** The report covers:

- the purposes of classroom assessment;
- the uses which are made of assessment results;
- the problems of 'high stakes' assessment;
- assessment methodology: kinds of evidence; the translation of evidence into judgement; the maintenance of assessment quality and consistency;
- the national assessment system of England compared with those of other countries, both within and outside the UK;
- assessing the quality of assessment systems in terms of *validity*, *reliability*, *impact* and *resources*;
- discussion of current national assessment arrangements in England in light of the above.

Purposes and systems of assessment

Assessment is a process of making judgements about pupils' learning; it involves decisions about what evidence to use and how to collect, interpret and report the judgements. The assessment of pupils' learning has two main purposes. One is to help that learning; this is referred to as **assessment for learning (AfL)** or **formative assessment**. The other is to summarise and report on what has been learned, referred to as **assessment of learning** or **summative assessment**. Whilst formative assessment has the sole purpose of helping pupils' learning, summative assessment is carried out for various school and external purposes, including those of statutory national assessment.

However, an assessment system is more than the procedures that are used for assessing individual pupils for different purposes; it also includes how standards of pupil performance are monitored at regional and national levels and how data on the performance of pupils is used in the accountability of teachers and schools. The parts of a system interact with each other. When, as in the system in England, the summative assessment results for groups of pupils are used for accountability, which includes setting targets for pupils' performance, and for monitoring national standards for pupils of certain ages and stages, year on year. These uses give the results of this assessment 'high stakes', since meeting or not meeting targets has important consequences for those held to be responsible. This affects not only teachers' own assessment practices but also the curriculum and methods of teaching and learning.

Different assessment systems

In England, a combination of national testing and teachers' judgments is used for summative assessment and national test results for classes and schools are used to evaluate the performance of schools, local authorities and the country as a whole. Examination of different assessment systems in other countries shows that, whilst England may be similar to many states of the USA in requiring high

stakes testing, in other countries - including the rest of the UK - steps are taken to avoid high stakes use of pupil summative assessment at the primary school level.

In France, Sweden and New Zealand, individual summative assessment is based on assessment by teachers; external standardised tests are not used for this purpose. This does not mean that there are no tests. In France and New Zealand, teachers are required to administer tests to all pupils at the start of certain years to help in their planning. In France, New Zealand and Scotland, teachers can use items from banks of tests and tasks to check their judgments of pupils' work for summative purposes. Tests are also used in these countries for the purpose of monitoring regional and national standards, but in all cases these are quite separate from tests used by teachers. These monitoring surveys involve a relatively small sample of pupils on each occasion and are designed to provide a wide sample of the subject domains assessed. They have low stakes because they cannot be used for school evaluation. In France, a sample of the tests given at the beginning of the year in the primary and first year of lower secondary school is also collected. But this is an anonymous sample, and the results cannot be used to report on the performance of individual schools.

Evaluating assessment systems

The pros and cons of different methods of conducting assessment for its different purposes and uses are identified by considering the evidence in relation to four key desirable properties of any assessment: **validity, reliability, positive impact** and **reasonable demand on resources**.

All the systems considered encourage, either implicitly or explicitly, the formative use of assessment, as happens in England, where AfL has begun to feature in the government's Primary Strategy, introduced in 2003. How far these measure help to produce the changes in teaching and classroom assessment practices that AfL requires is strongly influenced by other features of the assessment system.

The strongest impact on classroom practice comes from the way in which summative assessment for external purposes is carried out and the uses to which the resulting information is put. In England there are both national tests and assessment by teachers, but only the results of the tests are used for setting targets. In the interests of fairness, measures of pupil achievement needed for this use ought to be highly reliable. In practice, when tests are used, efforts to increase reliability result in the tests being restricted to those learning outcomes where performance can be most easily marked as correct or incorrect. This tends to exclude learning outcomes such as application of concepts, reasoning, understanding and attitudes to learning itself. Thus the validity of what is assessed is compromised by the requirement of high reliability, for the test does not sample the full range of important learning outcomes. In contrast, the validity of summative assessment *by teachers* is potentially high since they can gather evidence from the full range of learning activities, which cover all the goals.

The assumption that tests are necessarily more reliable than teachers' assessment is, in any event, not justified. Regardless of the consistency of individual test items, the fact that a test has to be limited to a small sample of possible items means that the test as a whole is a rather poor measure for any individual pupil. It is likely that a different selection of items would produce a different result. It is estimated that for the end of key stage tests in England this means that as many as one third of pupils may be given the wrong 'level'. Only an increase in length of test beyond anything that is practicable would materially change this situation. Thus there are limits to how accurate the results of tests can be. On the other hand, there are several ways of raising the reliability of teachers' assessment. The examples of practice in various countries show that the most commonly used are group moderation and the use of special tests or tasks that have been tried out and calibrated for teachers to use to check their judgements.

In relation to impact, there is considerable research evidence that high stakes tests put teachers under pressure to increase scores, which they do by teaching to the tests, giving multiple practice tests and coaching pupils in how to answer test questions. There is also firm evidence that this results in considerable stress for pupils. Other known consequences are the de-motivation of lower achieving pupils and, for all pupils, a view of learning as product rather than process. The proposal by the DfES of the introduction of single-level tests and the use of the results to set schools 'progress targets' is bound to increase these negative impacts of testing.

In relation to resources used in assessment, the cost of end of key stage tests in terms of teachers' and pupils' time is of considerable concern. Estimates based on surveys conducted in 2003 indicate that it is not the time spent on administering national tests, but the preparation for them that is most demanding. The extra time spent by teachers when external summative assessment is based on tests over that required for all other assessment activities is about 100 hours per year in Y2, and 165 hours, or about 5 weeks (at 33 hours per week). Estimates for pupil time spent on assessment suggest that practising and taking tests occupies the equivalent of about nine days a year in Y5 and 13 in Y6. This is time that teachers and pupils could use in other ways.

Teachers' summative assessment also requires time, of course, particularly for quality assurance procedures. There is compensation for this time, however, in the benefits to learning and teaching. Using teachers' judgments means that the evidence gathered in on-going activities can be used to help learning as well as providing the latest and best evidence that is reviewed in terms of reporting criteria at those time when this is required. Thus summative assessment by teachers can be in synergy with formative assessment.

Conclusions

After examining the research evidence Primary Review Research Report 3/4 concludes that:

- **The current system of assessment in England provides information of only low dependability whilst having some negative impacts on teaching and learning.**
 - **Evidence of changes in standards over the years does not support the claim that testing 'drives up standards'.**
 - **Alternative systems need to be considered. One in which summative assessment is based on teachers' judgements would provide information that is more valid than tests and at least as reliable, but it would be necessary to avoid high stakes being attached to the results by not using them for purposes other than reporting on individual pupils.**
 - **Accountability should be based on evaluation of a school's input and process variables and resources, and not only on pupil achievements.**
 - **For national monitoring, a regular sample survey, using a large bank of items, would give far more information than is provided by results of individual pupils who have all taken the same test.**
-

FURTHER INFORMATION

The report on which this briefing is based: Harlen, W. (2007) *The Quality of Learning: assessment alternatives for primary education*. (Primary Review Research Survey 3/4), Cambridge: University of Cambridge Faculty of Education. ISBN 978-1-906478-03-2.

The report is available at www.primaryreview.org.uk and is one of 32 Primary Review interim reports. Two of these deal with the opinion-gathering strands of the Review's evidence base. The remainder report on the thirty surveys of published research which the Review has commissioned from its 70 academic consultants. The reports are being published now both to increase public understanding of primary education and to stimulate debate during the period leading up to the publication of the Review's final report in late 2008.

The Primary Review was launched in October 2006 as a wide-ranging independent enquiry into the condition and future of primary education in England. It is supported by Esmée Fairbairn Foundation, based at the University of Cambridge Faculty of Education and directed by Professor Robin Alexander.

The Review has ten themes and four strands of evidence (submissions, community and national soundings, surveys of published research, and searches of official data). The report summarised in this briefing relates to the **Research Survey** strand and the theme **Curriculum and Assessment**.

Enquiries: The Administrator, The Primary Review, Faculty of Education, 184 Hills Road, Cambridge, CB2 8PQ. Phone: 01223 767523.

Email: enquiries@primaryreview.org.uk . Website: www.primaryreview.org.uk.

Press enquiries: richard@margrave.co.uk (Richard Margrave, Communications Director).

For the text of the full report, and for other interim reports in this series, go to www.primaryreview.org.uk/Publications/Interimreports .

This briefing, and the report which it summarises, have been commissioned as evidence to the Primary Review. The analysis and opinions they contain are the authors' own.